

REPORT DOCUMENTATION PAGE					Form Approved OMB No. 0704-0188	
<p>The public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing the burden, to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.</p> <p><b>PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.</b></p>						
1. REPORT DATE (DD-MM-YYYY) 11-03-2015		2. REPORT TYPE Final		3. DATES COVERED (From - To) 28-03-2013 – 27-03-2015		
4. TITLE AND SUBTITLE  Trend Monitoring and Forecasting				5a. CONTRACT NUMBER FA2386-12-1-4039		
				5b. GRANT NUMBER Grant AOARD-124039		
				5c. PROGRAM ELEMENT NUMBER 61102F		
6. AUTHOR(S)  Byeong Ho Kang				5d. PROJECT NUMBER		
				5e. TASK NUMBER		
				5f. WORK UNIT NUMBER		
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) School of Computing University of Tasmania Private Bag 87 Hobart TAS 7001 Australia				8. PERFORMING ORGANIZATION REPORT NUMBER  N/A		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)  AOARD UNIT 45002 APO AP 96338-5002				10. SPONSOR/MONITOR'S ACRONYM(S)  AFRL/AFOSR/IOA(AOARD)		
				11. SPONSOR/MONITOR'S REPORT NUMBER(S) AOARD-124039		
12. DISTRIBUTION/AVAILABILITY STATEMENT  Distribution Code A: Approved for public release, distribution is unlimited.						
13. SUPPLEMENTARY NOTES						
14. ABSTRACT <p>Recently, almost all web services, including Twitter, Google, Internet News, and Wikipedia, analyze their user created social data and detect the most popular terms that are discussed and searched within their community. The popular terms that are detected are published in the list, called 'Trending Topic' list. Awareness and utilization of trending topics plays a crucial role in various fields, including marketing, politics, and economics. In this three-year project, we monitor and analyze the trending topics in different online communities and provide a smart service by utilizing them. In this project, we achieved the following aims: 1) identifying the relevance of trending topic to a target domain, 2) predicting the popularity trends of trending topic, and 3) predicting the diffusion trends of trending topics among different online communities. With sponsorship from the US Air Force Office of Scientific Research (Contracts/Grant) FA2386-12-1-4039 and Grant), our research has allowed us to characterize and analyze the concept of trending topics in online communities, and develop the smart service using them. In this process, we have established a foundational literature on this topic.</p>						
15. SUBJECT TERMS Smart service, Trending topics, Topic diffusion, Trend monitoring, Trend prediction, Online community, Semantic analysis						
16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON	
a. REPORT	b. ABSTRACT	c. THIS PAGE			Hiroshi Motoda, Ph. D.	
U	U	U	SAR	14	19b. TELEPHONE NUMBER (Include area code) +81-42-511-2011	

Report Documentation Page		Form Approved OMB No. 0704-0188
Public reporting burden for the collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden, to Washington Headquarters Services, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to a penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number.		
1. REPORT DATE <b>25 MAR 2015</b>	2. REPORT TYPE <b>Final</b>	3. DATES COVERED <b>28-03-2013 to 27-03-2015</b>
4. TITLE AND SUBTITLE <b>Trend Monitoring and Forecasting</b>		5a. CONTRACT NUMBER <b>FA2386-12-1-4039</b>
		5b. GRANT NUMBER
		5c. PROGRAM ELEMENT NUMBER
6. AUTHOR(S) <b>Byeong Ho Kang</b>		5d. PROJECT NUMBER
		5e. TASK NUMBER
		5f. WORK UNIT NUMBER
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) <b>School of Computing,University of Tasmania ,Private Bag 87 Hobart TAS 7001,Australia,NA,NA</b>		8. PERFORMING ORGANIZATION REPORT NUMBER <b>N/A</b>
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) <b>AOARD, UNIT 45002, APO, AP, 96338-5002</b>		10. SPONSOR/MONITOR'S ACRONYM(S) <b>AFRL/AFOSR/IOA(AOARD)</b>
		11. SPONSOR/MONITOR'S REPORT NUMBER(S) <b>AOARD-124039</b>
12. DISTRIBUTION/AVAILABILITY STATEMENT <b>Approved for public release; distribution unlimited</b>		
13. SUPPLEMENTARY NOTES		
14. ABSTRACT <b>Recently, almost all web services, including Twitter, Google, Internet News, and Wikipedia, analyze their user created social data and detect the most popular terms that are discussed and searched within their community. The popular terms that are detected are published in the list, called ???Trending Topic??? list. Awareness and utilization of trending topics plays a crucial role in various fields, including marketing, politics, and economics. In this three-year project, we monitor and analyze the trending topics in different online communities and provide a smart service by utilizing them. In this project, we achieved the following aims: 1) identifying the relevance of trending topic to a target domain, 2) predicting the popularity trends of trending topic, and 3) predicting the diffusion trends of trending topics among different online communities. With sponsorship from the US Air Force Office of Scientific Research (Contracts/Grant) FA2386-12-1-4039 and Grant), our research has allowed us to characterize and analyze the concept of trending topics in online communities, and develop the smart service using them. In this process, we have established a foundational literature on this topic.</b>		
15. SUBJECT TERMS <b>Smart service, Trending topics, Topic diffusion, Trend monitoring, Trend prediction, Online community, Semantic analysis</b>		

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT <b>Same as Report (SAR)</b>	18. NUMBER OF PAGES <b>14</b>	19a. NAME OF RESPONSIBLE PERSON
a. REPORT <b>unclassified</b>	b. ABSTRACT <b>unclassified</b>	c. THIS PAGE <b>unclassified</b>			

**“Trend Monitoring and Forecasting”**

**Date 03/11/2015**

**Name of Principal Investigators (PI and Co-PIs):**

- e-mail address : Byeong.Kang@utas.edu.au
- Institution : University of Tasmania
- Mailing Address : Private Bag 87 Hobart TAS 7001, Australia
- Phone : 03) 6226 2919
- Fax :
- 

Period of Performance: 03/14/2012 – 03/13/2015

**Abstract:**

Recently, almost all web services, including Twitter, Google, Internet News, and Wikipedia, analyze their user created social data and detect the most popular terms that are discussed and searched within their community. The popular terms that are detected are published in the list, called ‘Trending Topic’ list. Awareness and utilization of trending topics plays a crucial role in various fields, including marketing, politics, and economics. In this three-year project, we monitor and analyze the trending topics in different online communities and provide a smart service by utilizing them. In this project, we achieved the following aims: 1) identifying the relevance of trending topic to a target domain, 2) predicting the popularity trends of trending topic, and 3) predicting the diffusion trends of trending topics among different online communities.

With sponsorship from the US Air Force Office of Scientific Research (Contracts/Grant) FA2386-12-1-4039 and Grant), our research has allowed us to characterize and analyze the concept of trending topics in online communities, and develop the smart service using them. In this process, we have established a foundational literature on this topic.

## **Introduction:**

### **Background**

By using different types of web-based services, such as search engines, social media, and Internet news aggregation sites, internet users can share and search information throughout the world. These services have caused a huge information-sharing paradigm shift by accumulating the unprecedented amount of social data. This large amount of user created social data is like an untapped vein of gold in the 21st century. Many information providers analyze their social data and detect the most popular terms that are discussed and searched within their community. The detected popular terms are called 'Trending Topic' (Aiello, Petkos et al. 2013). Trending topics have been provided by various companies, including Google, Yahoo, Baidu, and Twitter, for more than 5 years.

Trending Topics are estimated to reflect the real-world issues from the people's point of view. For example, Kwak et al. (Kwak, Lee et al. 2010) indicated that over 85% of trending topics in Twitter are related to breaking news headlines, and the related tweets of each trending topic provides more detailed information of people's opinions. Being able to recognize and utilize the trending topics, people are currently most interested in online web communities, may lead to opportunities for analyzing the market share in almost every industry and research fields, including marketing, politics, and economics.

In this project, we focused on monitoring, analyzing trending topics, and providing smart services based on trending topics for this three-year project. The aims of the project are classified as follows:

#### **AIM 1 Identifying the relevance of trending topic to a target**

The first aim was to develop the personalized relevance identification system that displays the relevance of trending topics to a target domain, an individual or organization. To accomplish this aim, we first collected trending topics from Trending Topics service, such as Google Trends, Twitter Trending Topics, and Google News. Then, we set up an electronic document management system as a target domain that includes all knowledge and activities having to do with a target object. Finally, we identified the relevance of trending topics to a target domain by applying the Term Frequency Inverse Document Frequency (TFIDF).

#### **AIM 2 Predicting the popularity trends of trending topics**

The second aim was to determine the feature that affects the popularity trends of trending topics and to build the model for predicting the popularity trends of trending topics. The popularity rank trends change dynamically; it may increase, fall or remain steady. To achieve this aim, we first analyzed the patterns of popularity trends of trending topics, and found the features that affect the popularity changes. Based on the features, we built the prediction model.

#### **AIM 3 Predicting the diffusion trends of trending topics among different communities**

The final aim was to develop the model for predicting the diffusion trends (scale and range) of trending topics that determine how the trending topics in one community diffused to other online communities. For this aim, we monitored online trending topics.

## **Experiment:**

### **Data Collection for Experiment**

For the experiment, we collected trending topic terms from Trending Topics Service, including Google Trends, Twitter Trending Topics, and Google News, and then extracted the related real-time articles (news articles or tweet postings) of trending topics. We crawled the data for two years (from 30<sup>th</sup> June, 2012 to 30<sup>th</sup> June, 2014). Web API was used for the data collection.

#### **I. Identifying the relevance of trending topic to a target**

In the experiments for the first aim, we used 5-months trending topics data in Google trends. In 5 months data, there are 17559 unique topics, and 46800 topics in total. In order to use trending topic as a proper dataset, it is crucial to disambiguate the exact meaning of trending topics. This is because trending topics service provides only the topic terms, such as short phrases, keywords, or hash tags with no detailed description. In order to extract the ambiguity, we collect real-time articles (news articles and tweet postings) that contains the certain trending topic term, and extract the related keywords that represents the representative keywords of the trending topic. We applied Term Frequency weighting (the most successful approach, according to the human evaluation (Han, Chung et al. 2014)) for extracting the representative keywords of a trending topic.

Then, it is necessary to identify the target domain in order to calculate the relevance of trending topic

to it. The target domain for this experiment is the combination of different countries' food blogs, which contains 22933 web documents, and has four continent categories (e.g. Asia), 14 area categories (e.g. East Asia) and 26 country categories.

For measuring the relevance of a trending topic to a target domain, we calculated the relevance weight of each document (in the target domain) to each set of a trending topic (trending topic term + extracted related keywords). In order to calculate the relevance, we applied Term Frequency Inverse Document Frequency (TFIDF) (Han and Kang 2012).

## II. Predicting the popularity trends of trending topics

In order to achieve this goal, we used two year of trending topics data from Twitter Trending Topics. Before we explain our experiment for the second aim, it is necessary to define the meaning of popularity trends. 'Trending Topics' list shows the top 10 trending topics in descending order of popularity. The lower the rank the higher the popularity, the higher the rank the lower the popularity. Based on the rank of a trending topic, it is possible to recognize the degree of current popularity of that topic. Hence, we focused on building a model for predicting the popularity rank trends of trending topic in order to achieve this aim.

Before we built the model for predicting the popularity trends of trending topics, we found some interesting patterns of popularity change patterns as seen in figure 1 and figure 2. For both figure 1 and 2, it shows the popularity rank pattern of U.S. twitter trending topics; x-axis indicates the lifetime of a specific trending topic and y-axis represents the ranking pattern of a trending topic (from rank 1 to 10). The first interesting pattern we discovered, as illustrated in figure 1, is the steady moment from the popularity change patterns. From the data analysis, 82% of the patterns have the steady moment around the midnight (20:00 – 02:00) in U.S. time. During the steady moment, the popularity rank does not increase or decrease. The second interesting pattern in figure 2 was detected from the trending topics that are related to the big events of celebrities or athletes. For example, if a celebrity was killed or hospitalized or athletes have a big match, the popularity rank of trending topic related to that event is always high (around rank 1). As you can see the figure 2, 'rick ross', 'moammar gadhafi' and 'heavy d' are very popular celebrities who were killed or hospitalized, and 'drew bree's', 'ryan braun', 'jorge posada' are the athletes who had a big match with the opposing team during that period.

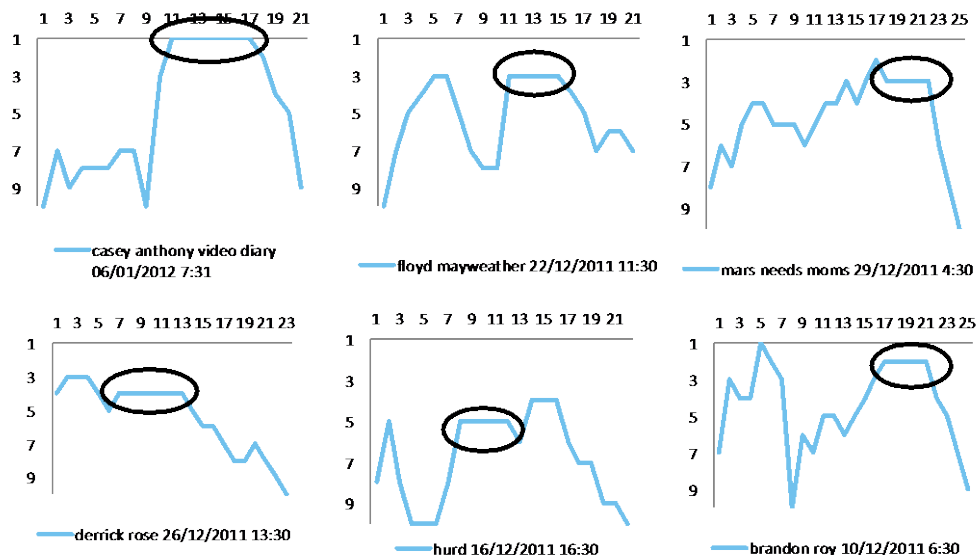
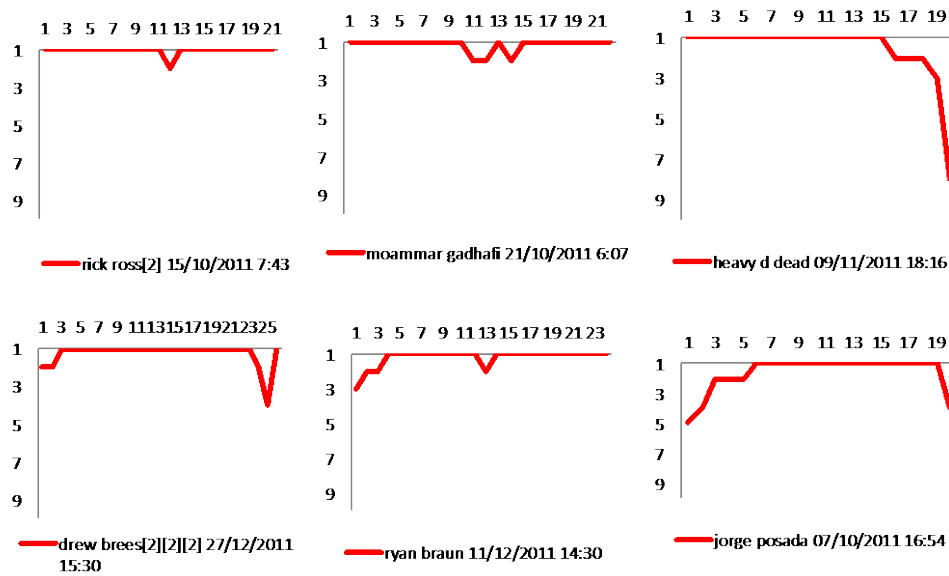


Figure 1 Steady moment from the trending topic popularity change patterns



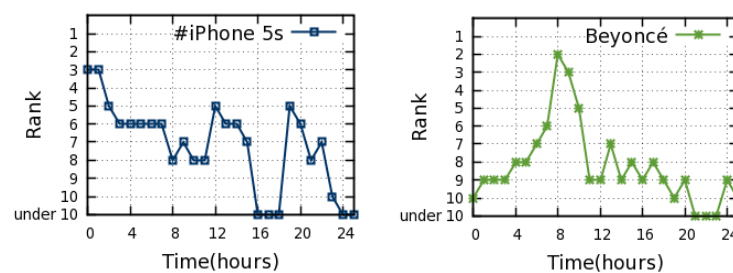
**Figure 2 trending topics popularity change pattern with celebrities**

For now, we understand that the popularity rank of trending topics represents the people's interests change, and the hourly ranking change can be classified into three categories: up, down, and unchanged. Therefore, we focused on answering "how can we predict the trends of trending topics' rank change (up, down, and unchanged) in the next hour?"

In order to solve this problem, we proposed a temporal modeling framework using historical rank and additional influential features. First, there were two main issues to solve when we used the historical ranking data for our model: missing ranking handling and window size selection.

#### First issue: Missing ranking handling

In twitter trending topics, it displays the top 10 trending topics (from rank1 to rank10) of the moment. In other words, if the topic disappears from the 'Trending Topics' list, it is impossible to recognize the exact ranking, whether the topic is ranked in 11<sup>th</sup> or 50<sup>th</sup>. Figure 3 shows the example of topic disappearance and reappearance from 'Trending Topics' list. Based on our analysis, almost 70% of trending topics tend to disappear and reappear to 'Trending Topics' list.



**Figure 3 The example of topic disappearance and reappearance from 'Trending Topic' list**

In order to deal with the missing ranking handling, we applied four successful missing value handling approach: 1) dummy variable control, 2) expectation maximization, 3) mean substitution, and 4) deletion.

#### Second Issue: Window size selection

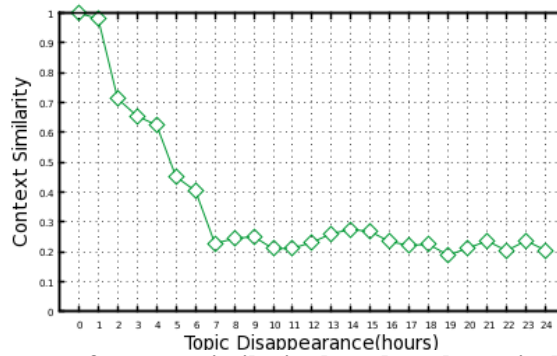
It is crucial to select the optimal window size in order to achieve time-series forecasting. We analyze the actual trending topic ranking data on U.S. Twitter, and the result shows that the same topic terms are sometimes referring to different events, and this normally occurs when the time length of the topic disappearance exceeds a certain time.

For example, there were two tragic events related to Malaysian Airlines in March and July, 2014; first event was flight disappearance, and second was flight bombing. As you can see in table 1, each extracted representative keywords in different time represent different events. Based on this data analysis result, we assumed that the length of topic disappearance time would affect to the event of a specific trending topic.

**Table 1 The example of same trending topic with different events**

Topic	Collected Date	Extracted Representative Keywords
'#MalaysiaAirlines'	2014/03/08	missing, flight, Malaysian, MH370, passenger, disappear, crash, pray, crew, lost, ocean, fail, safety, loss, airplane
'#MalaysiaAirlines'	2014/07/17	shot, down, missile, incident, kill, crash, attack, another, flight, victims, Malaysian, report, 259, explode

In order to select the optimal window size, we proposed the approach to identify the minimum length of topic disappearance that has different contexts by comparing the context similarity in two time-points. In detail, we firstly collected the trending topic and extracted the 15 (fifteen) related terms using term frequency (TF), and then calculated the context similarity of a specific trending topic at two different time-points (before-and-after the topic disappearance). Figure 4 shows the average of content similarity weight (1-exactly same / 0-completely different) based on the length of continuous topic disappearance time in U.S. trending topic. The context similarity is very low (0.2) when the topic continuously disappeared for over 7 hours.



**Figure 4 The average of content similarity based on the topic disappearance time**

#### **Additional feature**

In addition to the historical rank pattern, we used several features to improve the performance of our prediction model. We used semantic topic (same as the topic feature in aim 3) and starting time as additional features for the prediction model.

### **III. Predicting the diffusion trends of trending topics among different communities**

As mentioned earlier, trending topics show the popular issues among users in certain community (e.g. users in a certain web service or users in a certain country). Trending topics in one community can be different from others since the users in the community may discuss different topics from other communities. Surprisingly, we found that some trending topics are diffused among multiple communities. The third aim was to develop a model for predicting the diffusion trends (scale and range) of trending topics among different online communities; scale represents the number of communities that a trending topic diffuses, and range determines the depth of diffusion chain of a trending topic.

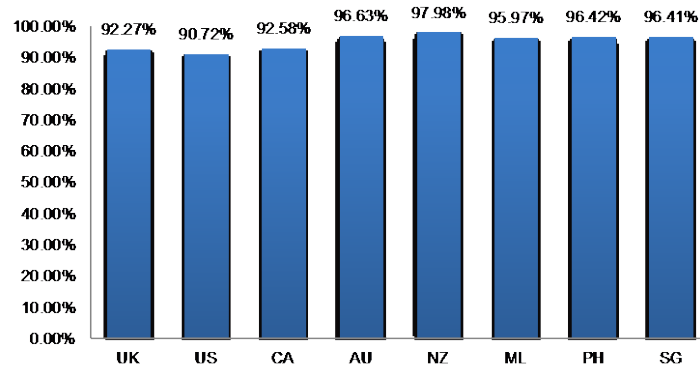
In this experiment, we tested the proposed model via two types of online trending topics diffusion.

1. Country-based trending topics diffusion prediction: We focused on predicting how a trending topic diffuses across multiple countries. We used twitter trending topics from 8 country communities (U.S., U.K., Australia, New Zealand, Canada, Malaysia, Philippine, and Singapore).



2. Web service-based trending topics diffusion prediction: We focused on predicting how a trending topic diffuses across web services. For this experiment, we used trending topics data from U.S. based Google Trends, Twitter Trending Topics, and Google News Top Stories.

The report shows the data analysis result for country-based diffusion trends but we will provide the prediction result of both country-based and web service-based trending topic diffusion in the Result and Discussion section.



**Figure 5 Percentage of trending topics diffused**

We found that over 90% of trending topics for each country appeared in different countries' trending topics list. For example, 92.27% of trending topics in UK appeared in at least one other country (only 7.73% of trending topic in UK appeared solely in UK). It represents that the trending topics are shared in not only one but also multiple countries. Therefore, predicting the diffusion trends of trending topics is a reasonable issue to solve.

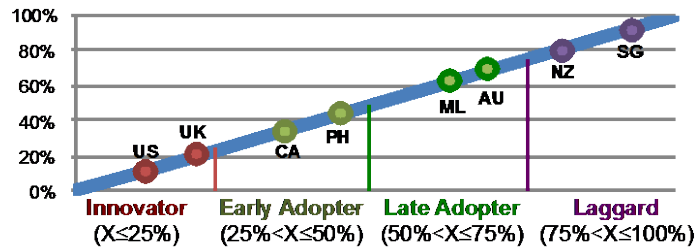
In order to predict the diffusion trends (scale and range) of trending topics, we applied the following four features in our prediction model.

a. Community Innovation Feature

The feature describes an innovation level of community of trending topics. It shows the level of the community adopt the trending topic. There are four types of innovation levels: 1) Innovator: Communities that start diffusing the trending topics, 2) Early Adopter: Communities that adopt the diffused trending topics in the early stage, 3) Late Adopter: Communities that adopt the diffused trending topics after the average participant, and 4) Laggards: Communities that are the last to adopt the diffused trending topic.

The way we classify the community innovation feature can be seen from figure 6 (community innovation level feature) – the graph shows the innovation level of each country. For example, for U.K. and U.S. it is the Innovator, and Canada (CA) and Philippines (PH) it is the Early Adopter.

Y-axis shows the percentage of share sectors (market share – the percentage of people who know the specific trending topics). X-axis represents the time. We classified the innovation levels based on the average percentage of time that a country spent on adopting the trending topics. For example, in average, topics are trending in the U.S online community when only 15% of English-speaking country communities adopts the topic.



**Figure 6 Community Innovation Level Feature**

b. Context Feature

This feature represents a context pattern of trending topics. We used three categories, including breaking news, meme, and commemorative day, on the context patterns. Table 2 shows the example pattern of classifying context pattern feature. Based on this context pattern, we created the rule and used over 20 rules to classify the trending topics using context patterns. For example, we used the following rule to find the 'meme'. If the trending topic contains '#' AND 'subject+verb', then trending topic is 'Meme'.

**Table 2 Context Feature Classification**

Patterns		Example	Category
Be + Noun		Nothing was the same	Meme
Verb + Noun		#HowToAskSomeoneOnADate	Meme
Person pronoun + Noun		#ILoveEXO	Meme
Possessive adjectives + Noun		#lookhowmanypeopleshowedupatyou rbirthdayliam	Meme
Noun only	Commemorative days	Christmas	Commemorative
	No commemorative days	Galaxy	News

In Table 3, you can see over 85% of trending topics are talking about the news, which is matched with the results from Kwak et al [1]. They mentioned that around 80% of trending topics are related to the title of breaking news.

**Table 3 Distribution of Context feature**

Categories	Percentage(%)
Commemorative	4%
Meme	9%
News	87%

c. Topic Feature

This feature represents the semantic topic of the Trending Topics. We classify the trending topics using NY Time topic classification service. The service provides nine (9) topic categories as follows:

- 1) Sports: trending topics that describe the sports games, athletes' names, and matching sports' name.
- 2) Entertainment: trending topics that describe celebrities, art and cultures, travel, movies, books, and theater

- 3) Politics: trending topics that describe politics' names, and parties.
- 4) Business: trending topics that are related to economy, business, career and workspace field
- 5) World issue: trending topics that are related to the world issue (affecting the world)
- 6) Technology: trending topics that describe technology, science, autos, and cars
- 7) Fashion: trending topics that describe home, lifestyle-leisure, and service-shopping
- 8) Obituaries: trending topics that describe crime, law, unrest, conflicts, war, disaster, and accidents
- 9) Health: trending topics that are related to health-related news, flu, and infectious disease

If we put a trending topic and the collected time into NY Times API, the service provides the category of articles, which contains the search query. We used the category based on the real-time article. The way to filter the real-time article is by checking the published time and seeing whether it matches with the trending topic collected time (trending topic collection time – 1 hour = the range of article published time).

Based on the data analysis result using country-based twitter trending topics, we found that 80% of trending topics are classified in the following topic categories, including entertainment, sports and politics (Table 4). It represents that most people are interested in the issues/events of entertainment, sports, or politics.

**Table 4 Distribution of Topic feature**

Topic Category	Percentage
Entertainment	42%
Sports	28%
Politics	10%
Fashion	6%
World issue	5%
Obituaries	4%
Health	2%
Business	2%
Technology	1%

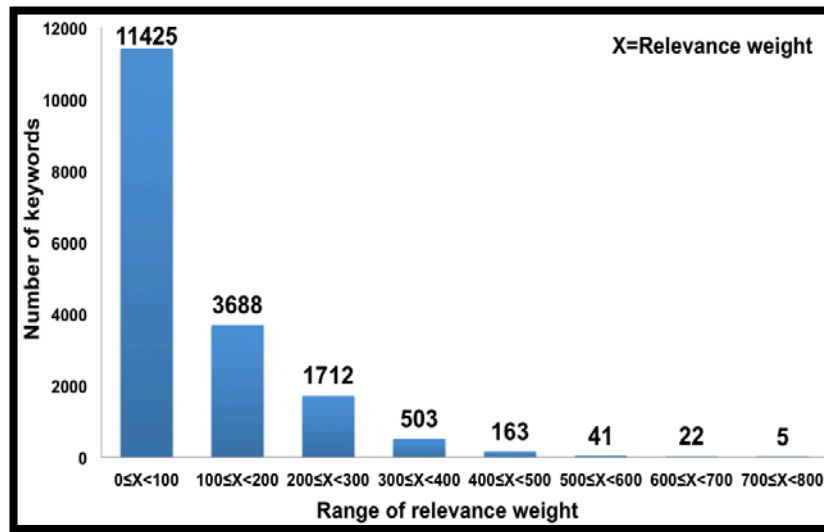
**d. Rank Feature**

Each trending topic has a popularity ranking (Rank 1 to 10). The ranking is changing in real-time. We used the ranking of a trending topic when it was initiated/started from a certain country. We show the percentage of initial ranking of trending topics.

**Results and Discussion:** Describe significant experimental and/or theoretical research advances or findings and their significance to the field and what work may be performed in the future as a follow-on project. Fellow researchers will be interested to know what impact this research has on your particular field of science.

**1) Identifying the relevance of trending topic to a target**

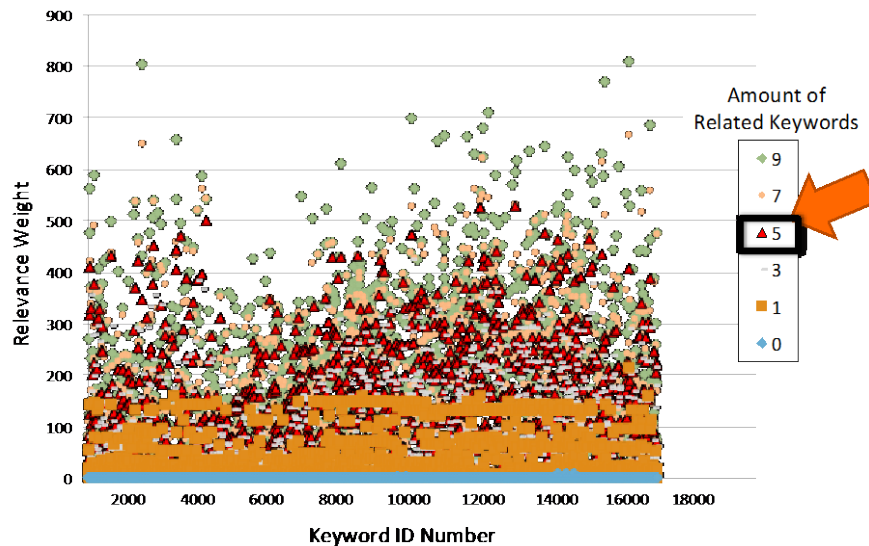
The first aim, relevance identification of trending topic to a target domain, is evaluated by using Google Trends trending topics (as trending topic data) and the combination of food blog (as target domain). Figure 7 shows the distribution of relevance weight of Google Trends trending topics to a target domain. The proposed system is able to clarify which topic is highly/lowly relevant to a target domain (Han and Chung 2012).



**Figure 7 Relevance Weights Distribution**

As mentioned before, we extracted related keywords in order to specify the exact meaning of trending topics. In this experiment, we would like to show the reason why it is necessary to extract several related keywords. We extracted ten related keywords for each Google Trends Trending Topic, and calculated their relevance weights to a target domain. Figure 8 represents the change of relevance weights with different number of related keywords.

Based on the figure, if we did not obtain any related keyword, the relevance weights are almost zero, illustrated as the blue line at the bottom of the figure. In this case, there may be some difficulty in defining which trending topic is highly related to a target object. However, if we extracted at least one related keyword, you can clearly see the big difference. This justifies why we need to extract the related keywords.



**Figure 8 Relevance weight based on the number of related keywords**

## 2) Predicting the popularity trends of trending topics

To accomplish the second aim, we built the temporal prediction framework using historical data and additional feature (semantic topic feature and time feature). The first result in Table 5 shows the popularity rank prediction accuracy of U.S. trending topic using only historical rank data. As

mentioned before, there are two main issues to solve: missing value (rank) handling and window size selection (# of instance). We applied and compared four missing value-handling approaches (**Dummy** - Zero, **EM**- Lowest+1, **Mean**, and **Deletion**). Based on our proposed window selection approach, the optimal size was 7 for U.S Trending Topic data.

**Table 5 Popularity Rank Prediction Accuracy of U.S. Trending Topics**

	# of instances	Missing Value	NB	NN	SVM	C4.5
(1)	5	Zero(0)	79.71%	88.20%	79.91%	88.74%
(2)	5	Lowest+1	80.11%	88.92%	80.82%	89.85%
(3)	5	Mean	75.10%	86.56%	77.29%	87.49%
(4)	5	Deletion	75.91%	85.42%	77.52%	85.74%
(5)	7	Zero(0)	83.91%	93.56%	85.36%	93.08%
(6)	7	Lowest+1	83.03%	93.68%	86.04%	94.01%
(7)	7	Mean	80.23%	91.06%	83.22%	92.91%
(8)	7	Deletion	82.93%	92.76%	83.93%	90.10%
(9)	9	Zero(0)	83.88%	92.53%	85.31%	93.00%
(10)	9	Lowest+1	83.00%	92.54%	85.61%	93.88%
(11)	9	Mean	80.34%	91.40%	83.29%	92.14%
(12)	9	Deletion	82.91%	90.92%	83.91%	90.11%

As can be seen in table 5, we found that using EM approach was the most successful approach for missing rank handling. Moreover, the proposed window size selection is working successfully. Surprisingly, rather than using complex features, we used historical ranking pattern and machine learning techniques, which achieved a successful result (94.01%).

After finishing the evaluation with historical data, we evaluated how much the performance can be improved with further features. We extracted the topic and time feature (as mentioned in the experiment part) and checked whether the performance had improved. However, with more features, there was only a one-percent increase (Table 6). However, it would be very difficult to perfectly predict the rank (100% accuracy). This is because popularity change is not based on algorithmic factors but has an irregularly changing nature.

**Table 6 Popularity rank prediction accuracy with different features**

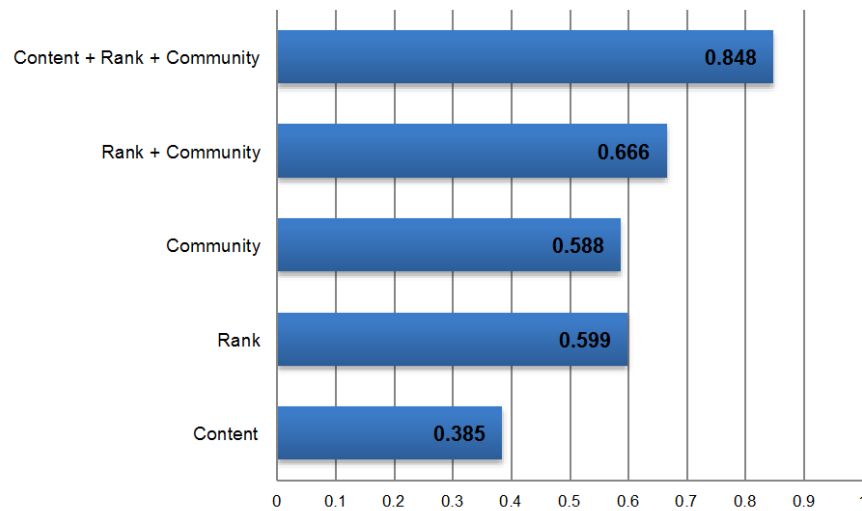
	HR	HR +Topic	HR+Topic+Time
Accuracy	94.01%	94.98%	95.01%

### 3) Predicting the diffusion trends of trending topics among different communities

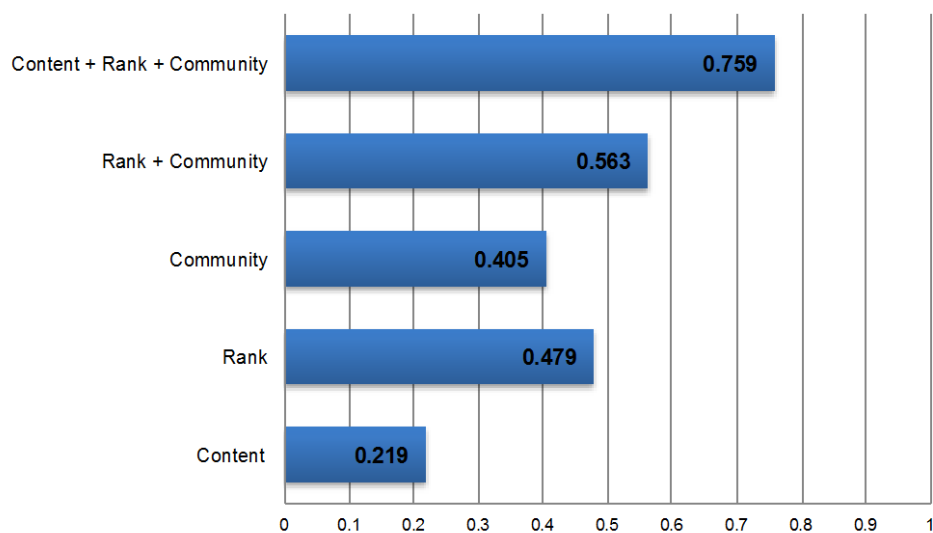
The two following figures, Figure 9 and 10, show the prediction accuracy of country-based trending topic diffusion trends (scale and range). We applied four features (community innovation level feature, context pattern feature, topic feature, and rank feature) into our prediction model. The model is learned by five different machine-learning techniques: Naïve Bayes, Neural Network, Support Vector Machine, Ridor, and C4.5 decision tree. Based on the results, prediction model learned by C4.5 decision tree achieved the highest prediction accuracy. The below Figure 9 and 10 are the results of the prediction accuracy with C4.5 decision tree.

As you can see in Figure 9, when we only use the content feature (topic feature and context feature), the accuracy result is lower than the others, which just reach 0.385 (scale) and 0.219 (range). However, by only using community innovation feature or ranking feature, the accuracy results almost reach 0.6 (scale) and 0.5 (range). When combining ranking feature and country feature, the prediction accuracy is increased. When it used all three features at the same time, it achieved the highest prediction

accuracy in both scale and range.



**Figure 9 Scale prediction Accuracy in country-based diffusion**



**Figure 10 Range Prediction Accuracy in country-based diffusion**

We applied the same prediction model for web service-based trending topics diffusion prediction. The prediction accuracy for web service-based trending topics is as follows: 75.01% (Scale) and 64.8% (Range). For both types of diffusion trends, country-based and web service-based, our proposed model performs better in scale than range. We can assume that the features in the proposed model are much suitable for scale prediction.

Compared to traditional social data applied diffusion prediction model, our proposed prediction model works successfully. However, it would be useful to discover further additional features that can improve prediction performance. Then, it can be used in trending topic diffusion prediction in other domains.

## Reference

Aiello, L. M., et al. (2013). "Sensing trending topics in Twitter." Multimedia, IEEE Transactions on **15**(6).

Han, S. C. and H. Chung (2012). Social issue gives you an opportunity: Discovering the personalised relevance of social issues. Knowledge Management and Acquisition for Intelligent Systems, Springer Berlin Heidelberg: 272-284.

Han, S. C., et al. (2014). Twitter Trending Topics Meaning Disambiguation. Knowledge Management and Acquisition for Smart Systems and Services, Springer: 126-137.

Han, S. C. and B. H. Kang (2012). Identifying the relevance of Social Issues to a Target. Web Services (ICWS), 2012 IEEE 19th International Conference on, IEEE.

Kwak, H., et al. (2010). What is Twitter, a social network or a news media? Proceedings of the 19th international conference on World wide web, ACM.

**List of Publications and Significant Collaborations that resulted from your AOARD supported project:** In standard format showing authors, title, journal, issue, pages, and date, for each category list the following:

a) papers published in peer-reviewed journals,

1. Han, Soyeon Caren, Hee-Geun Yoon, Byeong Ho Kang, and Seong-Bae Park. "Using MCRDR based Agile approach for expert system development." *Computing* 96, no. 9 (2014): 897-908.

b) papers published in peer-reviewed conference proceedings,

2. Han, Soyeon Caren, Hyunsuk Chung, Do Hyeong Kim, Sungyoung Lee, and Byeong Ho Kang. "Twitter Trending Topics Meaning Disambiguation." In *Knowledge Management and Acquisition for Smart Systems and Services*, pp. 126-137. Springer International Publishing, 2014.
3. Kang, Byeong Ho, Do Hyeong Kim, and Hyunsuk Chung. "What issue spread on the web: analyze the web trends." In *Proceedings of the 8th International Conference on Ubiquitous Information Management and Communication*, p. 24. ACM, 2014.
4. Han, Soyeon Caren, and Byeong Ho Kang. "Identifying the relevance of Social Issues to a Target." In *Web Services (ICWS), 2012 IEEE 19th International Conference on*, pp. 666-667. IEEE, 2012.
5. Han, Soyeon Caren, and Hyunsuk Chung. "Social issue gives you an opportunity: Discovering the personalised relevance of social issues." In *Knowledge Management and Acquisition for Intelligent Systems*, pp. 272-284. Springer Berlin Heidelberg, 2012.
6. Kim, Yang Sok, Byeong Ho Kang, Seung Hwan Ryu, Paul Compton, Soyeon Caren Han, and Tim Menzies. "Crowd-sourced knowledge bases." In *Knowledge Management and Acquisition for Intelligent Systems*, pp. 258-271. Springer Berlin Heidelberg, 2012.
7. Han, Soyeon Caren, Hyunsuk Chung, and Byeong Ho Kang. "It is time to prepare for the future: forecasting social trends." In *Computer Applications for Database, Education, and Ubiquitous Computing*, pp. 325-331. Springer Berlin Heidelberg, 2012.

d) Conference presentation without papers

8. Han, Soyeon Caren and Byeong Ho Kang. "Trending Topics Disambiguation", *AsiaPacific-Korea Conference 2014*, Sydney, Australia, November 20-22, 2014
9. Han, Soyeon Caren, Can Xie, and Byeong Ho Kang. "Trending Topics Diffusion Prediction", *AsiaPacific-Korea Conference 2014*, Sydney, Australia, November 20-22, 2014
10. Byeong Ho Kang. "Trending Topic Lifecycle Prediction", *AsiaPacific-Korea Conference 2012*, Singapore, December, 2012
11. Byeong Ho Kang, "Trending Topic Rank Analysis", *Korean Academy of Scientists and Engineers in Australasia Conference 2012*, Sydney, December, 2012

e) Manuscripts submitted but not yet published

12. Han, Soyeon Caren, Yulu Liang, Hyunsuk Chung, Hyejin Kim, and Byeong Ho Kang. "Chinese Trending Search Terms Popularity Rank Prediction", *Information Technology and Management*, (2015)

**Attachments:** Publications a), b) and c) listed above if possible.

**DD882:** As a separate document, please complete and sign the inventions disclosure form.

**Important Note:** If the work has been adequately described in refereed publications, submit an abstract as described above and refer the reader to your above List of Publications for details. If a full report needs to be written, then submission of a final report that is very similar to a full length journal article will be sufficient in most cases. This document may be as long or as short as needed to give a fair account of the work performed during the period of performance. There will be variations depending on the scope of the work. As such, there is no length or formatting constraints for the final report. Keep in mind the amount of funding you received relative to the amount of effort you put into the report. For example, do not submit a \$300k report for \$50k worth of funding; likewise, do not submit a \$50k report for \$300k worth of funding. Include as many charts and figures as required to explain the work.